

Weekly Report

2012.07.02-2012.07.08

黄芯芯

本周工作:

1. DataV.js 组件: 为了准备周末在阿里技术分享会的展示, 这周对自己开发的组件主要做了以下几项工作:
 - 1) 测试了散点矩阵在不同浏览器中的性能, 在 IE 下效率真的很低, 因此考虑降级, 在 IE 浏览器上不做交互。
 - 2) 根据浅夏的需求, 大费周章地实现一个交互, 深切体会到斐然师兄所讲的“需求总是不断变化”的难处。修改接口, 让用户选择类别。使用淘宝的实际数据。
 - 3) 周末去阿里技术分享会帮忙, 发现还是有挺多人对可视化感兴趣, 不过大多数也是停留在觉得“这个界面挺好看”的层面上。
2. 淘宝标签项目: 这星期和逢博、江华学长讨论了标签项目, 主要从“分析标签数据”、“马赛克图方法”、“Jing Yang 文章方法”以及“Treemap 方法”四个方面进行了讨论:
 - 1) 马赛克图方法是针对统计上的 Contingency Table 所做的方法, 所用的数据是一些交叉类别的频率数据, 而且对于表中的每个 Cell 都有一个相应的矩形来表示, 得到的可视化结果也是 $m \times n$ 型的(所以才叫做马赛克图)。而我们所有的标签数据并不是交叉类别的, 也不是统计频率的, 所以不知道如何用马赛克图。如果忽略马赛克图的统计意义, 只是仿照马赛克图的“水平上分维度, 然后垂直上分维度, 然后再水平上分维度……”的展示形式, 出来的结果和之前做的 Treemap 结果应该是一样的。
 - 2) Jing Yang 文章的方法: Jing Yang 的方法需要用在多个用户上, 每个一级类目作为一个维度, 一级类目下的二级类目作为一个类别值, 上周已经催了做数据的团队, 但是好像接口还是没做好。另外我们也考虑了一些问题: 每个用户买过的一级类目不一样, 所以有的用户在某些以及类目下的值可能是空的; 如果取买的商品数最多(或热度最高)的二级类目作为一级类目下的值, 这个是否有意义。
 - 3) Treemap 方法: 之前用多个 Treemap 的方法对一个用户买过的多个类目进行了可视化, 结果是有了, 但是也说不出这个结果有什么意义。这周讨论是否可以将全部的一级类目二级类目先用 Treemap 的方法可视化出来, 因为对于所有用户来说, 这个类目关系是统一的。然后, 对于每个单个的用户, 在他买过的类目下用颜色编码(可以用颜色编码买过商品的数量等), 这样做出来的结果应该是一个很稀疏的 Treemap, 因为一个用户每个过的商品并不会很多。

下周计划:

1. 组件还需继续修改, IE 降级等工作。
2. 标签项目和陈老师讨论, 希望有一个明确的方向。
3. 浙一那边的项目可能会有任务。